



A Close Look at the Relationship between Multiple Choice Vocabulary Test and Integrative Cloze Test of Lexical Words in Iranian Context

Parviz Ajideh

Tabriz University, Tabriz-Iran

E-mail: parviz_333@yahoo.com

Rajab Esfandiari

Tabriz University, Tabriz- Iran

Abstract

In spite of various definitions provided for it, language proficiency has been always a difficult concept to define and realize. However the commonality of all the definitions for this illusive concept is that language tests should seek to test the learners' ability to use real-life language. The best type of test to show such ability is considered to be the integrative test and cloze test is in turn regarded to constitute nearly all the factors needed for language use ability. However, the greatest obstacle of cloze tests or in general pragmatic tests is their administrative and scoring constraints for a large number of testees. Discrete-point tests, as the easiest and most common type of tests used for valid national and international proficiency tests, have always been doubtfully questioned as to whether they indicate the learners' ability to use language in real-life situation, but due to their tangible shortcomings, no absolute answer can be provided. The study aims at shedding light at the idea of the extent to which the discrete items of vocabulary proficiency show the learners' vocabulary proficiency in the real world of language use. Hence, the study seeks to calculate the correlation between discrete-point and integrative language proficiency tests of vocabulary administered to 21 Iranian freshmen studying English as a Foreign Language.

Keywords: Proficiency, Discrete-point test, Multiple-choice test, Integrative test, Cloze test, Correlation, Expectancy grammar, Variance, Validity, Communicative competence

1. Introduction

Language proficiency is one of the most poorly defined concepts in the field of language testing, and proficiency tests have always been a point of inquiry in language testing during the past decades. However, what all testing specialists unanimously agree upon is the ability of language use required of the learners. Brière (1972) points out that the parameters of language proficiency are not easy to identify. Acknowledging the complexities involved in the concept of language proficiency, Brière states that the term proficiency may be defined as: the degree of competence or the capability in a given language demonstrated by an individual at a given point in time independent of a specific textbook, chapter in the book, or pedagogical method.

Farhady (1982) objects the idea by pointing out the ambiguities of Brière's definition and maintains that such a complicated definition could very well result in vague hypotheses about language proficiency and language proficiency tests. They could be vague with respect to unspecified terms such as competence, capability, demonstrated, and individual. The term competence could refer to linguistic, socio-cultural, or other types of competence. The term capability could refer to the ability of the learner to recognize, comprehend, or produce language elements (or a combination of them). Demonstration of knowledge could be in either written or the oral mode. Finally, the expression individual could refer to a language learner as listener, speaker, or both. These concepts should be clarified and their characteristics should be identified in order to develop explicit hypotheses.

Clark (1972) Concerning language proficiency as the language learner's ability Clark (1972) states that to use language for real-life purposes without regard to the manner in which that competence was acquired. Thus, in proficiency testing, the frame of reference . . . shifts from the classroom to the actual situation in which the language is used.

Apart from its different definitions provided by prominent activists of the field, proper administration of proficiency tests to assess skills required of the testees has also been a matter of concern. To do the assessment two different approaches to testing the learners' skills have been proposed: discrete-point tests and integrative tests. Discrete-point tests have been criticized for their low reliability and validity. However integrative tests (e.g. cloze tests) have their own special problems. The biggest of all is the limitations of administration and scoring for a large group of testees. So as a doubtful solution to the problem of massive assessment of the learners' language proficiency to handle the real-life use

of language, in this study, discrete-point tests are selected both to facilitate the administration problems of integrative tests and to show the extent to which they are correlated with them.

2. Literature Review

2.1 Overview of Language and Teaching Testing

Language testing is one of the major areas in applied linguistics. It is an integral part of the instrumental program and plays an important role in education. If we assume that the purpose of a test is to ascertain whether or to what extent the learner knows the language, obviously, fundamental to the preparation of valid tests of language proficiency will be a theoretical question of what it means to know a language. Corder (1975) states that our ability to do a good job of measuring the learner's knowledge of the language depends upon the learner's knowledge of the language depends on upon the adequacy of our theory about the language , our understanding of what is meant by knowledge of language.

Surely, to know a language does not mean knowing something about the language, and any method of teaching based on this assumption will fail to meet this discussion. Prior to the emergence of the communicative approach to language teaching, two theoretical approaches attempted to find appropriate answers to this question. One is the habit-skill approach that views language behavior as a chain of habit units, and other one is the rule-governed grammar approach that views language competence as the ability to generate noel utterances on the basis of a finite set of rules.

Thus the first approach assumes that knowing a language is a kind of habit-formation through conditioning and drill. For the holders of this view, language is either mimicry or analogy, and grammatical rules are merely descriptions of what is called habits. As Diller (1978) states, for them the human being is essentially a machines with a collection of habits which have been molded by outside world.

The second approach assumes that to know a language is to be able to create new sentences. Unlike the first group, proponents of this approach do not refuse to talk of mind, for this approach do not refuse to talk of mind, for them the mind has a creative role in learning a language. Relying on cognitive theory, those who hold this view believe that it is impossible to know a language without thinking in it.

Both of these approaches study a language as an abstract form apart from its important characteristic as a means of communication.

With the emergence of the communicative approach in language teaching, it was assumed that to know a language , in addition to the ability to manipulate linguistic structures, a foreign language learner must also acquire knowledge of the rules and conventions governing their use for communication.

The communicative teaching movement has made clear developments in language teaching, and the communicative needs of the general language learners are favored by most course designer, syllabus writers and English teachers. In recent years, a need has arisen to specify the aims of language learning more precisely, and teaching of ESP rather than general English is favored by most of the English Learners throughout the world.

Cheng (2004) maintains that beliefs about testing to follow beliefs about teaching and learning. Early theories of test performance, influenced by structuralist linguistics, saw knowledge of language as consisting of mastery of the features of the language as a system. This position was clearly articulated by Robert Lado in his book *Language Testing*, published in 1961. Testing focused on candidates' knowledge of grammatical system, of vocabulary, and of aspects of pronunciation. There was a tendency to atomize and decontextualize the knowledge to be tested, and to test aspects of knowledge in isolation. Thus, tests of grammar would be separate from tests of vocabulary. Material to be tested was presented with minimal context, for example in an isolated sentence. According to McNamara (2000) this practice of testing separate, individual points of knowledge, known as discrete point testing was reinforced by theory and practice within psychometrics, the emerging science of measurement of cognitive abilities.

Within a decade, the necessity of assessing the practical language skills of foreign students wishing to study at universities together with the need within the communicative movement in teaching for tests which measured productive capacities for language, led to a demand for language tests involved an integrated performance on the part of the language user. The discrete point tradition of testing was seen as focusing too exclusively on knowledge formal linguistic system for its own sake rather than on the way such knowledge is used to achieve communication. The new orientation resulted in the development of tests which integrated knowledge of relevant systematic features of language with an understanding of context. As a result, a distinction was drawn between discrete point tests and integrative tests such as speaking in oral interviews, the composing of whole written texts, and tests involving comprehension of extended discourse. The problem was that such integrative tests tended to score, requiring trained raters; and in any case were potentially unreliable.

Research carried out by Oller, in the 1970s seemed to offer a solution. Oller (1973) offered a new view of language and language use underpinning tests, focusing less on knowledge of language and more on the psycholinguistic processing involved in language use. He suggested Pragmatic tests involving two factors: the online processing of language in real

time, and mapping of linguistic with extalinguistic factors. Further he proposed what came to be known as the Unitary competence Hypothesis, that is, that performance on a whole range of tests depended on the same underlying capacity in the learner-the ability to integrate grammatical , lexical, contextual, and pragmatic knowledge in test performance. He argued that certain kinds of more efficient tests, particularly the cloze test measured the same kinds of skills as those tested in productive tests. It was argued that of a cloze test was an appropriate substitute for a test of productive skills because it required readers to integrate grammatical, lexical, contextual, and pragmatic knowledge in order to be able to supply the missing words. But further work showed that cloze tests on the whole seemed mostly to be measuring the same kids of things as discrete point tests of vocabulary, grammar.

Douglas (2004) believes that, historically, language testing trends and practices have followed the shifting sands of teaching methodology. For example, in the 1950s, an era of behaviorism and special attention to contrastive analysis, testing focused on specific language elements. In the 1970s and in1980s, communicative theories of language brought with them a more integrative view of testing in which specialists claimed "the whole of the communicative event was considerably greater than the sum of its linguistic elements" (Clark, 1983, p. 432) . Today, test designers are still challenged in their quest for more authentic, valid instruments that stimulate real world interaction.

2.2 Discrete-point and Integrative Testing

The historical perspective underscore two major approaches to language testing that were debated in the 1070s and early 1980s. These approaches still prevail today. Even if in mutated form: the choice between discrete –point and integrative testing methods. Discrete point tests are constructed on the assumptions that language can be broken down into its component parts and that those parts can be tested successfully. It was claimed that an overall language proficiency test, then, should sample all four skills and as many linguistic discrete points as possible.

Such an approach demanded a decontextualization that often confused the test-taker. So, as the profession emerged into an era of emphasizing communication. Authenticity, and context, new approaches were sought. Oller (1979) argued that language competence is a unified set of interacting abilities that cannot be tested separately. His claim was that communicative competence is so global and requires such interaction that it cannot be captured in additive testsof grammar, reading, vocabulary, and other discrete points of language.) Others (Czikó, 1982, and savignon,1982) soon followed in their support for integrative testing.

Proponents of integrative test methods soon centered their arguments on what became known as the unitary trait hypothesis, which suggested an indivisible view of language proficiency: that vocabulary, grammar, phonology, the four skills, and other discrete points of language could not be disentangled from each other in language performance. The unitary trait hypothesis contended that there is a general factor of language proficiency such that all the discrete points do not add up to that whole.

Others argued against the unitary trait position. Farhady (1982) found significant and widely varying differences in performance on an ESL proficiency test, depending on subjects' native country, major field of study, and graduate versus undergraduate status. Weir (1990) noted that integrative tests such as cloze only tell us about a candidate's linguistic competence. They do not tell us anything directly about a student's performance ability.

2.2.1 Multiple-choice Items as Discrete-point Tests

A number of books discuss the construction and administration of multiple-choice items. 'Language Testing' by Robert Lado (1961), Modern Language Testing: A Handbook by Rebecca Valette (1967), Testing English as a Second Language by David Harris (1969), Foreign Language Testing: Theory and Practice by John Clark (1972), Testing and Experimental Methods by J. P. B. Allen and Alan Davies (1977), Revision of Modern Language Testing by Valette (1977) are some of the important books which deal with the construction of discrete point tests.

Discrete point multiple-choice tests assess one skill at a time – listening, speaking, reading or writing. They assess only one aspect of the skill – i.e. productive versus receptive, oral versus visual, etc. They attempt to focus attention on one point of grammar at a time. Each test item is aimed at one element of a particular component of a grammar item. According to Lado (1961) within each skill, aspect and component, discrete items focus on precisely one and only one phoneme, morpheme, lexical item, grammatical rule or whatever the appropriate element may be. But some believe that the reliability of multiple-choice tests is a function of the number of responses per item. They found that reduction in the number of distractors tended to lower the test reliability. Spearman - Brown formula gave reasonable good predictions of the reduced reliability when distractors were eliminated at random. To do further investigation some started with four-response forms by systematically eliminating the least effective distractor. They found that in a test period of fixed time limit, a greater number of two response items would produce more reliable scores than a smaller number of three of four response items. According to language testing specialists the essential characteristics of the distractors of multiple-choice items is that they should be plausible to those who lack the knowledge or ability for which the item is testing. Hence a lot of care should be put into the selection of the distractors.

2.2.2 Pragmatics

Pragmatics is concerned with the relationship between linguistic contexts and extralinguistic contexts. In this connection Oller (1979) states,

Pragmatics is about how people communicate information about facts and feelings to other people, or how they merely express themselves and their feelings through the use of language for no particular audience, except possibly an omniscient god. (p,19)

Oller (1979) adds that quite often we know much more than what we actually express in words. We also leave a lot of it unsaid and we depend on the receiver to fill in what is unsaid and interpret our message. In normal use of language, no matter what level of language or mode of processing we think of, it is always possible to predict partially what will come next in any given sequence of elements. The elements may be sounds, syllables, words, phrases, sentences, paragraphs, or larger units of discourse. The mode of processing may be listening, speaking, reading, writing, or thinking, or some combination of these. In the meaningful use of language, some sort of pragmatic expectancy grammar must function in all cases. (p,25)

2.2.3 Expectancy Grammar

According to Oller (1976), the notion of an expectancy grammar characterizes the psychologically real system that governs the use of a language in an individual who knows that language. The characteristic of such an expectancy system helps in two ways: to explain why certain kinds of language tests apparently work as well as they do; and to device other effective testing procedures that take account of these salient characteristics of functional language proficiency.

A valid language test should press the learners' internalized expectancy system into action and must further challenge its limits of efficient functioning in order to discriminate among degrees of efficiency. A language test to be valid should meet the **pragmatic naturalness criteria**. A test is said to meet the pragmatic naturalness criteria when it invokes and challenges the efficiency of the learners' expectancy grammar by causing him to process temporal sequences in the language that can conform to normal contextual constraints and by requiring him to understand the systematic correspondences of linguistic and extralinguistic context.

2.2.4 Pragmatic Tests and Language Proficiency

According to Oller (1979) there are two aspects of language use: factive and emotive use, the first is applied to convey information about people, things, events, ideas and states of affairs and the second is used to convey our attitude about the factual information we want to convey.

Every time we use language, we use both the aspects of language. It is quite possible for people to agree on the factual information conveyed but differ on the attitude towards those facts. There are two major contexts of language use: first the linguistic context which refers to the verbal and gestural contexts of language; and second the extralinguistic context which refers to the states of affairs constituted by things, events, people, ideas, relationships, feelings, perceptions, memories and so forth. The objective aspect of linguistic context, the world of existing things, may be distinguished from the subjective aspect of extralinguistic context, the world of self-concept and inter-personal relationships. There are systematic correspondences between linguistic and extralinguistic contexts. Linguistic contexts are pragmatically mapped onto extralinguistic contexts, and vice versa.

3. Methodology

3.1 Research Question and Research Hypotheses

This study is aimed at answering the following research question:

3.1.1 Research question:

Is there a significant difference between the result of discrete point (multiple choice) item type test of vocabulary and integrative cloze test of lexical words?

In other words,

Can discrete-point test of vocabulary be used instead of integrative cloze test of lexical word test in massive assessment of the learners' language proficiency?

3.1.2 Research Hypotheses:

Null hypothesis: There is no significant correlation between these two kinds of tests.

Hypothesis: There is a significant correlation between these two kinds of tests.

3.2 Participants

The participants of the study consisted of two groups of young freshmen studying at Tabriz University. The age range of the participants of both groups varied between 19 and 25 with different first languages. Their sex was not a

controlled factor. The qualities of the two groups are considered to be homogeneous in terms of their proficiency due to the alphabetical arrangement criterion used for dividing them into two groups. The first group is just to take a non-standardized 50-item multiple-choice test of vocabulary for standardization procedure and is supposed to be exactly equal to second group. The second group who is to take the standardized multiple-choice test of vocabulary with the cloze test consists of 21 freshmen.

3.3 Procedure

This section deals with the selection procedure of the two to-be-administered tests. As for the multiple-choice test, first of all 50 multiple-choice tests of vocabulary were made for freshmen level of proficiency and administered to the first group. After administration the tests were primarily standardized through checking their item difficulty and item discrimination and rearranged according to their difficulty level. The outcome became a standardized 30-item multiple-choice test of vocabulary revised for the difficulty and discrimination power of its items. In other words, the test administered for primary revision included 50 items out of which 30 relatively standard items were selected for the standardized test.

The cloze test also included 30 blank items of lexical words which were selected subjectively according to the variable ratio method of item deletion. The scoring procedure adopted for the cloze test was the contextually appropriate word method. To calculate the correlation between the 30-item multiple-choice standardized test of vocabulary and the cloze test, the two tests were administered to the second group of freshmen.

3.4 Data analysis

In this part of the study some statistical procedures, data tabulation, display of graphs and interpretive statistics of the second group's test-taking will be explained. Some of them include the Mean, Variance, and Standard Deviation of the standardized multiple-choice test of vocabulary and the cloze test. Then the main statistical concern of the research (i.e. the correlation) will be discussed.

In the table 1 the second group's cloze test and multiple-choice test scores are shown. The scores of each individual are very near to each other. The approximate overlap between the two sets of scores is observable in figure 1.

Upon doing some descriptive statistical procedures the following data in table2 were obtained for the Mean, Variance, and Standard Deviation of the standardized multiple-choice test of vocabulary and the cloze test.

The sets of data in table 1 are delineated in figure 2.

The figure delineates well that mean, variance, and the standard deviation of the two sets of scores are near to each other. However to depend upon these descriptive figures is a premature judgment. So to base the study upon reliable data calculation procedures the appropriate process is to obtain the correlation between the two sets of scores. In the following we investigate the research hypotheses and the correlation between cloze test and multiple-choice test scores.

3.5 Result

The correlation between the standardized multiple-choice test of vocabulary and the cloze test we came to the figure .57 which is a relatively high correlation between two kinds of tests which are seemingly very different. Thus it is safe to reject the null-hypothesis and approve the research hypothesis that there is a significant correlation between discrete-point item test of vocabulary and integrative cloze test of lexical words.

4. Conclusions

According to what is obtained as the correlation between the relatively standardized multiple-choice vocabulary test scores and the cloze test scores of vocabulary the following conclusions can be made:

- 1). In testing the proficiency of a group of learners the overall result of the multiple-choice vocabulary test scores are very much like that of the cloze test scores.
- 2). According to the correlation worked out, multiple-choice tests of vocabulary could be a substitute for cloze tests of vocabulary in massive development of proficiency tests.

5. As shown by the results of the study it could be concluded that those who act better on discrete Point vocabulary also act better in cloze test of vocabulary.

4.1 Limitations of the Study

The limitations of the study were the following:

- 1). If the number of the participants was more than 21 participants the correlation would probably be strengthened.
- 2). The more the number of the items in both the discrete-point test and the cloze test the more the correlation of the between the two.

4.2 Pedagogical Implications

The study has some implications for test-makers, language teachers, and syllabus designers and may be for others who are concerned with language tests of proficiency, teaching, and developing materials for EFL or ESL students. However three of them are referred to here.

- 1). The first implication will be for test makers. In most cases test-makers could make tests to test testees' knowledge of language through separate points of language (e.g. grammatical components or vocabulary items) especially in occasions being short of time to assess the proficiency of a large group of testees.
- 2). The second implication will be for language teachers in EFL or ESL settings. It is not recommended to teach language through separate components of language. But according to the study teaching through exposing language learners to integrative samples of language can just increase the rate of learning but not change the route of learning and they will seemingly have approximately the same outcome. So inclusion of discrete points of language in teaching can also be helpful not only to enable the learners to perceive and produce extensive stretches of language but also to draw analytic attention of the learners' to constituting parts of language.
- 3). The third implication goes to syllabus designers. Recently there was a tendency towards looking upon language as a holistic entity as a reaction to common analytic approaches to language teaching of past decades. But current synthetic approaches allow syllabus designers to not only put emphasis on the discoursal level of language knowledge but also show secondary concerns for the parts of it. Thus it is upon syllabus designers to include both trends in the materials they develop. However they could prioritize the holistic view of language as the primary goal and the analytic view of language as the secondary one and not to totally dispense with the latter in hope of enabling the students to fully master language use.

5. Suggestions for Further Research

According to the limitations of the study and also because of being obliged to choose one of two forms of a factor involved in doing the study (e.g. scoring cloze tests either by contextually appropriate word method or by exact word method, deleting words either by fixed ratio method or by variable ratio method) there remain some other questions which could undergo further investigations. Below are some of the suggestions for either obviating the shortcomings of the present study or going through the unexplored dimensions of it.

- 1). The present study worked out the correlation between discrete-point tests of vocabulary and cloze test items of lexical words. To investigate new dimensions of the research question the correlation between discrete-point tests of grammar and cloze test items of both lexical and functional words could be calculated.
- 2). As mentioned before one of the shortcomings of the present study was the small number of testees who took part in the study. To strengthen the validity of the correlation, the number of the testees could be increased and doing so, its effect on the magnitude of the correlation could be certainly positive.
- 3). One of the other shortcomings of the study is the limited number of the test items used. To increase the validity of the obtained correlation the number of test items in both the multiple-choice of vocabulary and the cloze test could get increased to 50 or even 100 test items (however for the cloze test two or more texts could be used in order to have balanced distribution of blanks).
- 4). In the present study the participants were freshmen just having been accepted at the university. The amount of the correlation for other levels of proficiency could be obtained too.
- 5). In this study the participants' age and sex were taken for granted. Further investigation could be done to evaluate the proficiency of different age and sex groups.
- 6). In the present study to assess the testees' vocabulary knowledge only those items in the cloze test were deleted that were regarded as lexical items. So the procedure adopted for the deletion of words was inevitably a variable ratio method. For further research the fixed ratio method could be used to assess the effect of the change in deletion procedure on the magnitude of the correlation between the two kinds of tests.
- 7). For scoring the cloze tests the scoring procedure in this study was the contextually appropriate word method. To assess the effect of the change in the scoring procedure on the correlation magnitude between the two kinds of tests the exact word method of scoring could be adopted.

References

- Allen, J. P. B., & Davies, A. (Eds.). (1977). Testing and experimental methods. The Edinburgh Course in Applied Linguistics (Vol. 4). London: Oxford University Press.
- Brière, E. J. (1972). Are we really measuring proficiency with our foreign language tests? In H. B. Allen & R. N. Campbell (Eds.), Teaching English as a second language: A book of readings (2nd ed.), 1972. New York: McGraw-Hill Book Company.

- Cheng, L, Yoshinori Watanbase, Andy Curtis. (2004). Washback in Language Testing. Lawrence Erlbaum Association, Inc.
- Clark, J. L. D. (1972). Foreign language testing: Theory and practice. Philadelphia, PA: The Center for Curriculum Development.
- Clark John L.D. (1983). Language testing: Past and current status, directions for the future. Modern language journal, 67,431-443.
- Corder, S. Pit. (1975). Introducing Applied Linguistics. Penguin Education, Hazell Watson and Vincyltd.
- Diller, K.C.(1978). Generative Grammar, Structural Linguistics, and Language Teaching. Rowley, Mass: Newbury house.
- Douglas, H. Brown. (2004). Language Assessment Principle and Classroom Practice. Pearson Education, Inc.
- Farhady, H. (1982). Measures of Language Proficiency from the Learner's Perspective. TESOL Quarterly 16, 1.
- Harris, D. P. (1969). Testing English as a Second Language. New York: MacGraw Hill Book Company.
- Lado, R. (1961). Language testing: The construction and use of foreign language tests. New York: McGraw-Hill Book Company.
- Mcnamara, Tim. (2000). Language Testing. Oxford: Oxford University Press.
- Oller, J. W., Jr. (1973). Cloze tests of second language proficiency and what they measure. Language Learning 23:105-119
- Oller, J. W., Jr. (1976). Evidence of a general language proficiency factor: An expectancy grammar. Die Neuen Sprachen, 76, 165-74.
- Oller, J. W., Jr. (1979). Language tests at school: A pragmatic approach. London: Longman.
- Savignon, Sandara, J. (1982). Dictation as a measure of communicative competence in French as a second language. Language Learning, 32.
- Valette, R.M. (1967). Modern Language Testing: A Handbook (first edition), New York, Harcourt, Brace and World
- Valette, R.M. (1977). Modern Language Testing: A Handbook (second edition), New York, Harcourt, Brace and World.
- Weir, C.J. (1990). Communicative Language Testing, Prentice Hall, Englewood Cliffs, NJ.
- Zicko, Gray A. (1982). Improving the psychometric, criterion-referenced, and practical qualities of integrative language use. TESOL Quarterly, 16.

Appendices

Table 1. *Cloze Test and Multiple-choice Test Scores*

Student	Cloze Test Score	Multiple-choice Test
1	27	27
2	27	22
3	27	18
4	26	25
5	26	19
6	26	27
7	24	25
8	23	23
9	23	25
10	23	24
11	23	21
12	23	25
13	23	24
14	22	23
15	22	25
16	21	16
17	21	22
18	19	18
19	17	19
20	16	15
21	15	18

Table 2. Mean, Variance, and Standard Deviation of the standardized multiple-choice test of vocabulary and the cloze test

Test Type	Mean	Variance	Standard Deviation
Multiple-choice Test	21.95	12.84	3.58
Cloze Test	22.57	12.35	3.51

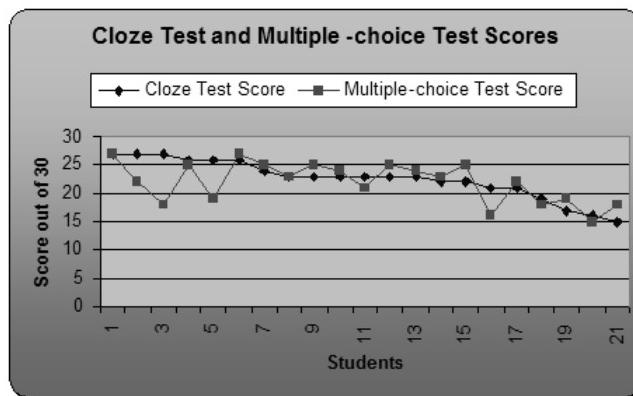


Figure 1. Cloze Test and Multiple-choice Test Scores

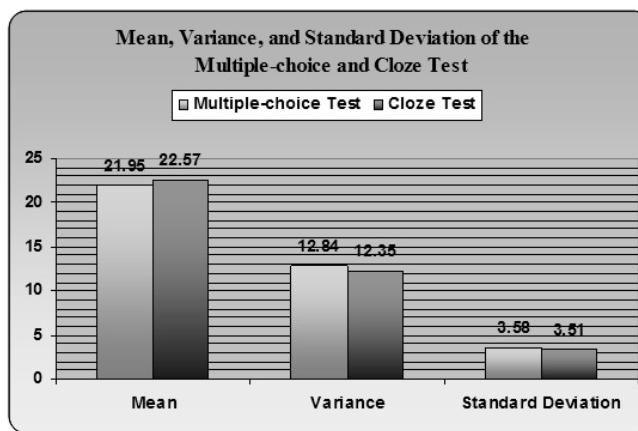


Figure 2. Mean, Variance, and Standard Deviation of the multiple-choice test of vocabulary and the cloze test